



Interaction of visual prior constraints

Pascal Mamassian^{a,*}, Michael S. Landy^b

^a Psychology Department, University of Glasgow, 58 Hillhead Street, Glasgow G12 8QB, Scotland, UK

^b Psychology Department and Center for Neural Science, New York University, New York, NY, USA

Received 3 March 2001; received in revised form 17 April 2001

Abstract

The visual system relies on two types of information to interpret a visual scene: the cues that can be extracted from the retinal images and prior constraints that are used to disambiguate the scene. Many studies have looked at how multiple visual cues are combined. We examined the interaction of multiple prior constraints. The particular constraints studied here are assumptions the observer makes concerning the location of the light source (for the shading cue to depth) and the orientation of a surface (for depth based on image contours). The reliability of each of the two cues was manipulated by changing the contrast of different parts of the stimuli. We developed a model based on elements of Bayesian decision theory that permitted us to track the weights applied to each of the prior constraints as a function of the cue reliabilities. The results provided evidence that prior constraints behave just like visual cues to depth: cues with more reliable information have higher weight attributed to their corresponding prior constraint. © 2001 Elsevier Science Ltd. All rights reserved.

Keywords: Visual constraints; Three-dimensional perception; Bayesian modelling; Shape from shading; Shape from contours

1. Introduction

One of the fundamental problems that the visual system faces is the ambiguity inherent in retinal images. An image could have been produced by an infinite number of objects of different shape, size, orientation and colour. However, not only are we rarely aware of these ambiguities, but a given image is likely to be interpreted the same way by different observers and by the same observer at different times.¹ The consistency with which ambiguous images are interpreted supports the idea that the visual system uses assumptions to help in image interpretation (Rock, 1983). In previous work (Mamassian & Landy, 1998), we proposed a methodological framework to characterise these assumptions, and applied this framework to the assumption that our viewpoint is above the objects we are observing. In this paper, we examine the way these assumptions interact.

The assumptions used by the visual system are known as *constraints* in the computational approach to vision (Marr, 1982). Within this framework, constraints are used to find unique solutions to ill-posed problems. For instance, assuming that objects are rigid is a powerful constraint that allows one to estimate the three-dimensional structure of a moving object (Ullman, 1979). In this now classical work, the constraints were combined with sensory data within a regularisation framework where a carefully chosen cost function was minimised (Poggio, Torre, & Koch, 1985). Although this framework has found many successful applications in computational vision (Horn, 1990), one major drawback for the understanding of natural vision is the ad hoc use of constraints that are often chosen for mathematical convenience (e.g. prior distributions or cost functions that result in a tractable minimisation problem).

Rather than considering visual constraints as merely a technique to render a problem well-posed, one can treat constraints on an equal basis with sensory data. The Bayesian framework provides an explicit way to optimally combine constraints and sensory data (Kersten, 1990). In fact, one can show that regularisation

* Corresponding author. Fax: +44-141-330-46061.

E-mail address: pascal@psy.gla.ac.uk (P. Mamassian).

¹ Ambiguous stimuli that lead to bi-stable percepts (seen differently by different observers and by the same observer at different times) are, in fact, rare. The stimuli used in this manuscript are among these.

models are just a special class of Bayesian models (Yuille & Bülthoff, 1996). The Bayesian framework has already proved useful in the interpretation of line drawings (Mamassian & Landy, 1998) and motion patterns (Hogervorst & Eagle, 1998; Weiss & Adelson, 1998).

The purpose of the present study is to use the Bayesian framework in scenes where multiple constraints interact to influence the percept. A classical example of such an interaction is when one views a concave mask of a human face illuminated from above (Gregory, 1980). The mask is consistent with two mutually exclusive interpretations: either a concave mask lit from above or a convex mask lit from below. On the one hand, we are used to viewing convex faces, but, on the other hand, we tend to assume that light comes from above the viewed object (e.g. Ramachandran, 1988). Looking at a concave mask lit from above therefore puts two constraints in conflict; for a self-consistent percept, one or the other constraint must be violated. In this case, the visual system is reluctant to abandon the face convexity constraint, thereby leading to an illusory percept of a normal, convex face lit from below.

The interaction of multiple constraints can be compared to the interaction of multiple sensory cues. In both cases, potentially conflicting sources of information are combined to produce a unique, stable percept. The ‘modified weak fusion model’ has recently been proposed for the interaction of multiple depth cues (Landy, Maloney, Johnston, & Young, 1995). One characteristic of the model is *dynamic re-weighting*, whereby the weight of each cue is based on the reliability of that cue relative to the reliabilities of other cues present in the scene. The reliability of each cue is assessed using ancillary measures such as velocity for a motion cue, or viewing distance for a binocular cue (depth estimated using motion and binocular disparity are likely to be less reliable for smaller velocities and larger viewing distances, respectively). In the present study, we shall see how dynamic re-weighting can also be applied to the interaction of multiple constraints.

In the rest of the paper, we report the results of an experiment that characterises how human observers combine two visual constraints. These constraints are our assumptions that light comes from above (Mamassian & Goutcher, 2001), and that our viewpoint is located above the object of regard (Mamassian & Landy, 1998), both of which are described in detail. We then develop a model inspired by Bayesian decision theory that accounts for the data. This model is used to determine the weights assigned to each constraint. We look at the variation of the weights as a function of characteristics of the stimuli. We conclude the paper with a discussion of the relevance of the Bayesian framework to model depth perception.

2. Experiment

The purpose of the psychophysical experiment was to determine how human adults deal with two visual constraints. We know from previous work that human observers rely on a priori assumptions when sensory information is ambiguous (Gregory, 1980; Rock, 1983). What happens when two constraints lead to inconsistent interpretations? Does the strongest constraint veto the weakest or do the two constraints interact? If the constraints interact, how are the weights attributed to each constraint?

To address these questions, we concentrate on two constraints: the assumptions that light comes from above and that one’s viewpoint is located above the scene. These two constraints play a role in the interpretation of three-dimensional shape from shading and parallel contours. For instance, the same shading patterns can be produced by a convex object lit from below and by a concave object lit from above. Assuming where the light source is can therefore disambiguate the shape of an object (e.g. Ramachandran, 1988). In previous work, we have shown that, by default, human observers assume that light is coming from above-left, with a bias to the left between 20° and 30° off the vertical (Mamassian & Goutcher, 2001; also see Sun & Perona, 1998). Similarly, parallel contours painted on the surface of an object provide a strong cue for shape, up to another convex–concave ambiguity (Stevens, 1981). This ambiguity can be resolved if one knows the general direction of the normals to the surface. In previous work, we have shown that human observers consistently interpret images containing parallel contours as surfaces with their normals pointing upwards (Mamassian & Landy, 1998). This bias for surface normals pointing upwards corresponds roughly to a preference for one’s viewpoint to be located above the object. In the remainder of the paper, we shall refer to the preference for the location of the directional light source as the *shading constraint* and the preference for the direction of the surface normal as the *contour constraint*. We describe here an experiment using stimuli for which both constraints could be used, and stimulus conditions that put the two constraints either in accord or in conflict. We then vary systematically the reliability of the depth cues corresponding to each constraint and observe how the reliability affects the degree to which one constraint dominates the other.

The sort of stimuli used is shown in Fig. 1A. It depicts a patch of surface that appears to have a series of raised ridges, running from top-left to bottom-right, crossed with dark surface contours. Observers generally perceive Fig. 1A to have the narrow regions bulging toward the observer, and the wider regions as the valleys, although the stimulus is ambiguous and can be perceived with the wide strips bulging. Why are the

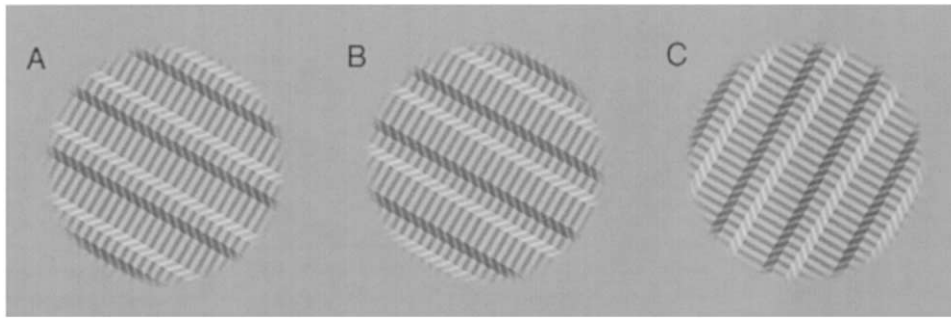


Fig. 1. Sample stimuli. (A) In this stimulus, most observers perceive narrow ridges and wide valleys, a percept consistent with the assumptions of a light source and viewpoint above the object. (B) Rotating the stimulus in (A) by 180° reverses the percept; now the narrow strips are seen indented and wide strips are bulging. (C) Rotating the stimulus in (A) by 90° results in one for which the two assumptions imply opposite interpretations. This stimulus is more ambiguous for most observers.

narrow strips usually perceived as bulging? First, consider the shading cue. Notice that the bevelled regions between the narrow and wide strips alternate in intensity (bright, then dark, then bright again). When the narrow strips are seen bulging, the intensities of these bevelled regions are consistent with matte shading with a light source located above the observer. Next, consider the depth cue of shape from surface contour. The constraint that the observer's viewpoint is above the object being viewed (Mamassian & Landy, 1998) implies that when a surface contour is convex-upward in the image, it is generally perceived as convex toward the observer. Thus, the painted surface contours in Fig. 1A running across the narrow and wide strips are also consistent with the narrow strips bulging.

Now, consider the surface shown in Fig. 1B. This is identical to Fig. 1A, except that it has been rotated in the image plane by 180°. Both constraints associated with the shading and contour cues are consistent with the wide strips being seen bulging. Indeed, that is how most observers perceive this stimulus.

Finally, consider the surface shown in Fig. 1C. This is identical to Fig. 1A, except that it has been rotated clockwise by 90°. The pattern of shading, coupled with the assumption that light comes from above, implies that the wide strips are bulging. The surface contours and their associated constraint imply the opposite. As expected, this stimulus is the most ambiguous of the three. In the experiment, the degree to which the shading and contour constraints dominate the percept in these cue-conflict stimuli was used to determine the relative weights used by observers for the two cue constraints.

2.1. Methods

2.1.1. Subjects

The observers were eight graduate students and post-doctoral fellows in the psychology department of the University of Glasgow. All had normal or corrected-to-normal eyesight.

2.1.2. Stimuli

The stimuli depicted embossed surfaces that were primarily planar, and had either narrow or wide strips bulging (Fig. 1). The strips appeared in relief because of the shaded edges of the strips (the edge facing the light was brighter than the edge in shadow), and because of parallel contours that ran perpendicularly to the orientation of the strips. The intensity along the edge of a strip was constant and chosen such that all shaded edges had the same contrast that we shall call the *shading contrast* C_S . Similarly, the contrast between the dark and bright parts along the parallel contours was constant and equal to the *contour contrast* C_C . These contrasts were thus

$$C_S = \frac{B_3 - B_2}{B_3 + B_2} = \frac{B_2 - B_1}{B_2 + B_1} = \frac{D_3 - D_2}{D_3 + D_2} = \frac{D_2 - D_1}{D_2 + D_1} \quad (1)$$

and

$$C_C = \frac{B_1 - D_1}{B_1 + D_1} = \frac{B_2 - D_2}{B_2 + D_2} = \frac{B_3 - D_3}{B_3 + D_3}, \quad (2)$$

where B_1 , B_2 and B_3 are the intensities along the bright parallel contours, and D_1 , D_2 and D_3 are the intensities along the dark parallel contours (Fig. 2). These are physically realisable stimuli that would result from painted, matte surfaces. C_S derives from the relative amounts of ambient and point source illumination. C_C is the contrast between the two surface reflectances corresponding to the bright and dark contours.

Three levels of C_S (0.05, 0.1 and 0.2) were combined with three levels of C_C (0.1, 0.2 and 0.4) leading to nine contrast conditions (Fig. 3). For all the stimuli shown in Fig. 3, when the contour running across the narrow strips is convex-upward in the image (as shown), the bright bevelled regions always lie on the left side of this ridge. To counter-balance the relative positions and effects of the shading and contour cues, two sets of stimulus figures were used: that illustrated in Fig. 3 and its mirror reflection (shown, e.g. in Figs. 1 and 2). Hence, there were 18 conditions (nine contrasts times

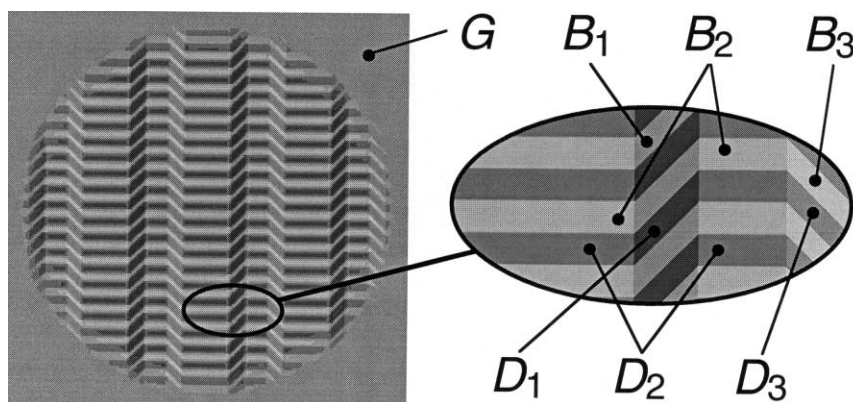


Fig. 2. Detail of the different intensities present in a given pattern. These intensities were determined using given values of shading and contour contrast.

two mirror reflections). For each condition, the corresponding stimulus was presented at each of 24 orientations in the frontal plane.

The background luminance G was set to mid-grey (37.2 cd/m^2). With the additional constraint that the mean of the six stimulus luminances equalled the background luminance, the strip luminances that satisfy Eqs. (1) and (2) can be computed for each pair of shading and contour contrast:

$$D_1 = G(1 - C_C)(1 - C_S)/(1 + C_S)$$

$$D_2 = G(1 - C_C)$$

$$D_3 = G(1 - C_C)(1 + C_S)/(1 - C_S)$$

$$B_1 = G(1 + C_C)(1 - C_S)/(1 + C_S)$$

$$B_2 = G(1 + C_C)$$

$$B_3 = G(1 + C_C)(1 + C_S)/(1 - C_S), \quad (3)$$

The stimulus was windowed by a disc whose edges gradually faded to the background. The diameter of the disc, and therefore the size of the stimulus, was 2° of visual angle. Within this disc, four periods of narrow and wide strips were visible. Wide strips were twice as wide as narrow ones. Parallel contours ran perpendicularly to the strips, alternating between dark and bright contours of equal width.

2.1.3. Materials

The experiment was run in a darkened room and the frame of the monitor was the only object that was still slightly visible. The stimuli were displayed on a 17-inch SONY Trinitron display driven by the Psyscope software (Cohen, MacWhinney, Flatt, & Provost, 1993) running on an Apple Power Macintosh computer. A chin rest restricted subjects' head movements. Viewing was monocular, alternating between the left and right eyes between blocks of trials. Viewing distance was 1 m.

2.1.4. Procedure

The task of the observer was to decide whether the initial percept was that of an embossed surface with 'narrow' or 'wide' strips bulging. They responded with the left and right arrow keys on the computer keyboard, using their left and right index fingers. Each observer ran four blocks of 432 trials (18 stimuli in each of 24 orientations). Each block took approximately 15 min to complete.

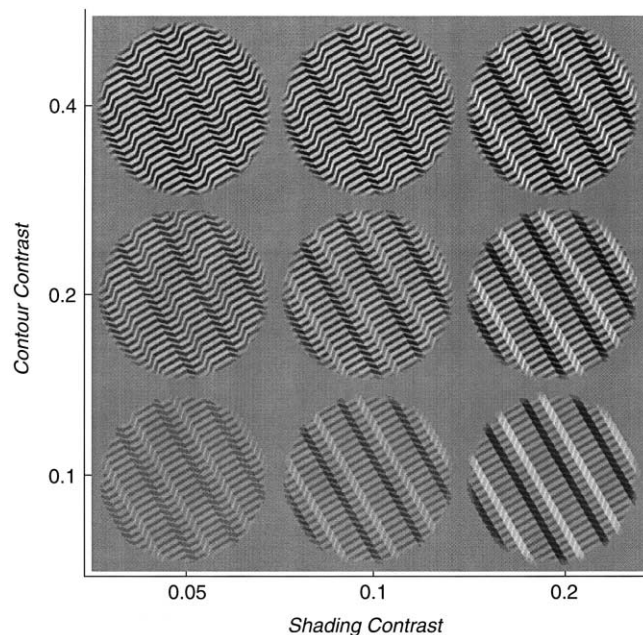


Fig. 3. The nine different stimuli obtained by crossing three levels of shading contrast and three levels of contour contrast. Nine other patterns were obtained by using the mirror reverse of these. Each of these images was presented in one of 24 orientations in the image plane (three of which are shown in Fig. 1).

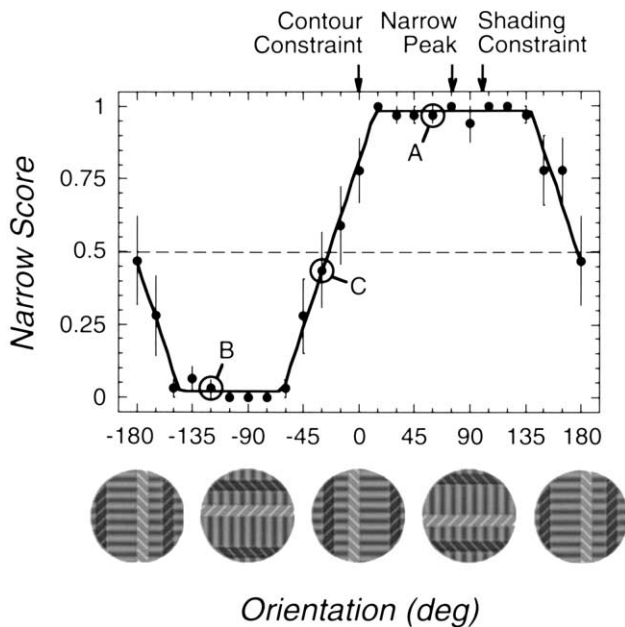


Fig. 4. Results averaged across eight observers for the pattern shown in Fig. 1 ($C_S = C_C = 0.2$). The abscissa shows the orientation of the stimulus in the image plane. The ordinate gives the proportion of times observers indicated that the narrow strips were perceived bulging (chance performance is 0.5). The data points corresponding to the three stimuli of Fig. 1 are circled. Error bars are standard errors of the mean showing variability across observers. The solid line is a piecewise linear curve fit to the data from which the peak of the narrow score is estimated (see text for details). Also indicated are the orientations that best indicate that the narrow strips are bulging according to the shading and contour constraints.

2.2. Results

The results for the pattern illustrated in Fig. 1 ($C_S = C_C = 0.2$) are shown in Fig. 4. The abscissa shows the orientation of the pattern in the frontal plane. The ordinate gives the *narrow score*: the proportion of times that a pattern was perceived with the narrow strips bulging. The narrow score was averaged across all subjects and was therefore based on 32 presentations (four repeated presentations for each of eight observers). The three orientations shown in Fig. 1 are highlighted and led to ceiling, floor, and chance performance, respectively. These results confirm our observation in the introduction that a given pattern can be perceived as either 'narrow' or 'wide' simply by rotating the pattern in the frontal plane. The error bars show the standard error of the mean across the eight observers' scores.

Also shown in Fig. 4 are the orientations corresponding to peak narrow scores predicted by the shading and contour constraints. The prediction for the shading constraint was obtained by assuming that the shading cue was most effective when the shading contours were oriented orthogonally to the preferred light direction. (We used a value of 10.7° to the left of vertical, taken

from the fit of the full model in Section 3.2.8). Similarly, the peak narrow score predicted for the contour constraint was obtained for the orientation of the figure that corresponded to the surface normals pointing precisely upwards. Note that the human data peak falls between the values predicted by the shading and contour constraints.

The results for the nine different combinations of shading and contour contrast are shown in Fig. 5. Each figure contains two plots, one for the pattern in Fig. 3 and the other for its mirror reflection. A systematic shift of the peak narrow score is seen in both curves as the relative contrast of the two depth cues changes from that favouring use of contours (upper-left plot) to that favouring shading (lower-right plot).

To better illustrate this trend, we automatically estimated the orientations leading to the peak narrow scores for each condition. Each curve was fit using a piecewise linear function composed of a range of floor performance, a linear increase, a range of ceiling performance, and finally a linear decrease. This function is characterised by the floor performance p (ceiling performance is set to $1 - p$), the rising and falling transitional orientation θ_R and θ_F (the orientations that lead to a narrow score of 0.5), and the slope s of the rising portion (the slope of the falling portion was set to $-s$). A minimum root-mean-square (RMS) error fit of this curve to the data of Fig. 4 is shown as the solid line. The orientation θ_P leading to a peak narrow score is then computed as the average of the rising and falling points of subjective equality (PSEs).

Fig. 6 shows the estimated peak orientations θ_P for the nine contrast combinations and for one of the two mirror-reversed figures (that corresponding to the right-hand curves in each panel of Fig. 5). The point corresponding to the curve of Fig. 4 is circled. The peak orientation increases as the shading contrast increases, gradually shifting the peak of the curve toward the orientation consistent with the shading constraint. Similarly, the peak orientation decreases as the contour contrast increases, shifting the peak toward the orientation consistent with the contour constraint. Similar effects were observed for the other mirror-reflected figure, for which the shading constraint predicts a peak near -90° .

2.3. Discussion

When two constraints are potentially relevant for the interpretation of ambiguous images, the human visual system uses both in a cooperative way. When the two constraints agree as to the best interpretation of the image, observers usually adopt that interpretation. When the two constraints imply opposite interpretations, the probability of choosing one or the other depends on the quality of the stimulus information for

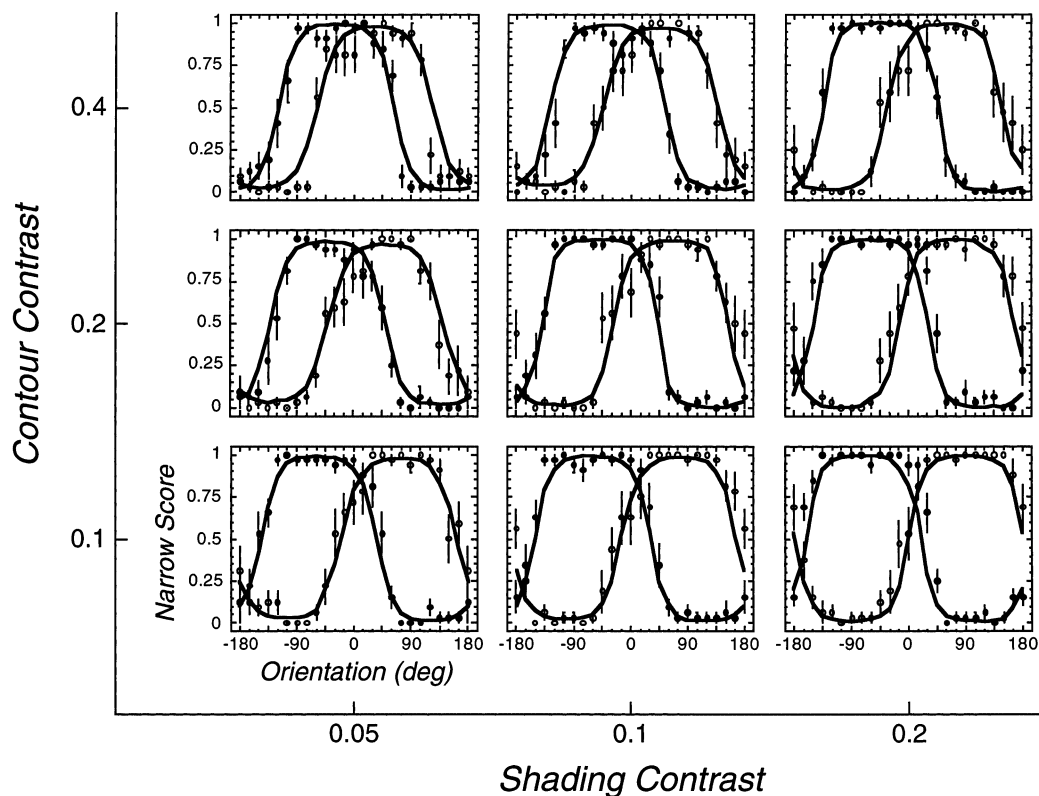


Fig. 5. Results averaged across eight observers for all 18 stimulus patterns. The solid and open symbols correspond to the two mirror-reversed figures for a given combination of shading and contour contrast. Error bars are standard errors of the means computed across observers. Solid curves are the results of the full model described in Section 3.2.

the cue corresponding to each constraint. It is as if each constraint is given a weight based on the stimulus information, and the relative weights determine the perceptual outcome. This is reminiscent of the dynamic allocation of weights for sensory cues to quantitative depth (Landy et al., 1995), where perceived depth is a linear combination of the rendered depth for each of the available depth cues, with the weights dependent on the respective cue reliabilities. In the psychophysical experiment just described, the reliability of the shading constraint was determined by the shading contrast, and likewise, the reliability of the contour constraint was determined by the contour contrast.

In the next section, we provide a method to determine quantitatively the weights assigned to each constraint. The method consists in developing a model of our psychophysical task and then fitting the model to the human performance data. This method will allow us to relate the shape of the prior distribution functions to the shading and contour contrasts.

3. Models

In this section, we derive models for the psychophysical experiment described in the preceding section. The

purpose of the models is to help us understand the principle by which the two constraints interact. The models are inspired by Bayesian models of visual perception (Knill & Richards, 1996; Maloney, 2001). Even though we shall use the terminology of Bayesian decision theory, our first model (the ‘full model’) differs significantly from traditional Bayesian models in ways that we shall highlight. We begin with a brief non-technical overview. We then describe the full model in

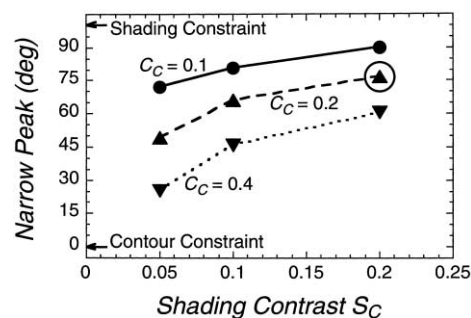


Fig. 6. Estimated peak orientations determined by averaging the rising and falling PSEs from piecewise linear fits to the right-hand datasets in each of the nine contrast combination conditions of Fig. 5. The stimulus corresponding to Fig. 4 is circled. Note that the peaks shift to larger orientations (i.e. toward the values consistent with the shading constraint) as the shading contrast S_C increases and as the contour contrast C_C decreases.

detail. We describe the fit of the full model to the experimental data to determine the weights of the prior constraints used for the various stimulus conditions (combinations of C_S and C_C), as well as fits of a more constrained model (the ‘nested model’). After some discussion, we describe a third model based on a different set of premises (the ‘internal noise model’).

3.1. Overview of the approach

The purpose of the model is to specify how sensory data are integrated with prior constraints so as to lead to the narrow scores observed in human performance. Similar to Bayesian models, it has three components: (i) a calculation of the likelihood function (the probability that the given sensory data would be observed given that the world is in state X); (ii) a description of the constraints in terms of prior distribution functions; and (iii) a decision rule (Yuille & Bülthoff, 1996). Bayesian models combine sensory data and prior constraints in an optimal way to determine the posterior distributions (the probability that the world is in state X given the sensory data and constraints). The decision rule links the posterior distributions to the observed performance data. Because the likelihood functions and the decision rules are mostly determined by the task, the Bayesian framework allows us to focus all our attention on the prior distributions.

The likelihood function can only be expressed at the cost of making explicit all the attributes of the scene that are likely to affect the appearance of the image. To this end, we will need to have models for the illumination, the surface shape and surface orientation. Along the way, we shall describe probability distribution functions for all the random variables that play a role in the model. For most of these random variables, we let their distributions be uniform on their respective domains, as a starting point, because we have no particular reason to believe the contrary.

A limited number of random variables will correspond to the prior constraints of the model. In these cases, the probability distribution functions will be non-uniform with one or two degrees of freedom to characterise the bias and the variance. The variance corresponds to the strength of the prior; as the variance approaches infinity, a prior distribution becomes increasingly uniform, that is, the prior has less and less effect. The mean of the prior can be used to model the preference induced by the constraint. For instance, in the model that follows, there is a distributional bias term to model the preferred direction of illumination (light comes from above, from above-left, etc.). The distributional parameters will then be fit to the psychophysical data. There will be three prior constraints: (i) a constraint that light is coming from one preferred direction (the shading constraint); (ii) a constraint on the

orientation of the surface (the contour constraint); and (iii) a general bias to respond ‘narrow’ rather than ‘wide’.

Finally, a decision rule should be selected. The decision rule determines the response given all the information available. In statistical decision theory, a number of common decision rules are employed such as maximum likelihood, maximum a posteriori and, more generally, maximum expected gain. However, each of these rules is deterministic. That is, for our experiment, for a given set of sensory data, each of these decision rules would either always respond ‘narrow’, or always respond ‘wide’. Human observers, on the other hand, often produce data values that lie anywhere between 0% and 100% ‘narrow’ responses. Thus, these decision rules cannot be a good model of human behaviour without recourse to additional processes to allow responses to be variable. In psychophysics with stimuli that are difficult to detect (e.g. contrast sensitivity measurements), variable responses are attributed to noise in the stimuli (photon noise) or noise sources in the visual system that corrupt internal detector responses. For easily visible, ambiguous depth stimuli, we prefer to adopt a different approach. We suggest that the variability in the response reflects the lack of confidence of choosing one particular interpretation over the others. In previous work, we used a non-committing rule where the probability of response is set to the posterior probability (Mamassian & Landy, 1998). Such a rule will be applied here as well, so that the model will respond ‘narrow’ on the same proportion of trials as the posterior probability of ‘narrow’.

3.2. The full model

We now detail the models for the illumination, surface shape and surface orientation. Where we have good reasons to believe that the observer might have preferences, we provide a model with parameters that are then fit to the human data. For all the other random variables, we assume by default that they are taken from uniform distributions on the domain where they are defined. We first start with a note about circular statistics, which will prove useful for the definition of prior probabilities.

3.2.1. Circular statistics

Circular statistics have to be used whenever probability distribution functions are periodic. In our experiment, the stimuli were presented at a random orientation in the frontal plane, so that the ‘narrow score’ that we try to model is periodic with period 2π . Many of the intervening variables used in the model are angles as well (slant and tilt for the surface orientation and illumination direction). Therefore, the Gaussian distribution, which is often used in modelling, has to be

replaced by a circular distribution. A standard choice for such a circular distribution is the von Mises distribution (Batschelet, 1981), which is defined as

$$M(\varphi, \kappa, \theta) = \frac{1}{2\pi I(0, \kappa)} e^{\kappa \cos(\varphi - \theta)}, \quad (4)$$

where θ is the bias, κ is the concentration parameter (the distribution is flat when $\kappa = 0$), φ is an angle ($\varphi \in [0, 2\pi]$), and I is the modified Bessel function. The variance of a von Mises distribution is inversely related to its concentration parameter.² Like the Gaussian distribution in linear statistics, the von Mises distribution is uni-modal and symmetric around its mode.

3.2.2. Illumination model

The illumination of a surface by a point light source defines patterns of shading that result from the variation of surface orientation relative to the lighting direction. Conversely, the shading on a surface provides a cue for surface shape (see Mamassian & Kersten (1996) for a review).

To account for the illumination conditions, we used the classical Phong model with ambient and Lambertian terms, but with no specular term (Foley, van Dam, Feiner, & Hughes, 1996). The ambient term corresponds to an illumination source that has a constant value throughout the scene independent of position and orientation, whereas the Lambertian term corresponds to the effect of a point light source for matte surfaces. If I_a and I_p denote the intensities of the ambient and point lights, respectively, the irradiance I at each point of the surface is given by

$$I = r(I_a + I_p \max\{\langle N, L \rangle, 0\}), \quad (5)$$

where N stands for the surface normal, L is a unit vector pointing in the direction of the point light source, and r is the reflectance (albedo) of the surface at the considered location. In this equation, $\langle x, y \rangle$ denotes the inner product between the vectors x and y . The ‘maximum’ function reflects the fact that attached shadow regions are completely occluded from the point light source.

With respect to our stimuli, the illumination model has six degrees of freedom: two degrees of freedom characterise the reflectance of the bright and dark surface contours (r_B and r_D), two more account for the intensities of the ambient and point light sources (I_a and I_p), and the last two determine the location of the point light source (slant σ_L and tilt τ_L).

We have no reason to believe that one albedo is more likely than another. Therefore, pending other evidence, the values of the reflectance for the bright and dark contours (r_B and r_D) will be taken from a uniform distribution on the interval $[0, r_{\max}]$. However, recall that the reflectance of the contours played a role in the psychophysical experiment only through the contour contrast C_C . In other words, the actual values of the reflectances do not matter, only the ratio between the bright and dark albedos is important. We can therefore set the maximum of the albedo range r_{\max} to an arbitrary value, and with no loss of generality, we set $r_{\max} = 1$.

Likewise, we have no reasons to assume any bias on the intensity of the light sources. The values of the intensity of the ambient and point light sources (I_a and I_p) were therefore taken from a uniform distribution on the interval $[0, I_{\max}]$. However, because the corresponding independent variable is again a contrast (S_C), the maximum of the intensity range I_{\max} is an arbitrary value and so, with no loss of generality, we set $I_{\max} = 1$.

The light direction will be described by two angles relative to the observer: the slant σ_L and the tilt τ_L . This parameterisation is tantamount to assuming that the point light source is at a large distance compared to the size of the surface so that light rays are parallel. The slant σ_L is the angle between the light direction and the viewing direction, and the tilt τ_L is the angle between the projection of the light direction in the frontal plane and the horizontal right direction. The slant σ_L characterises the illumination direction in terms of how far it is in front or back of the observer. We have no reason to assume any bias on the slant of the light direction. Therefore, the slant will be taken from a distribution that results in a uniform distribution across the sphere of possible illumination directions (at least when the distribution of τ_L is uniform):

$$p(\sigma_L) = \sin(\sigma_L). \quad (6)$$

The tilt τ_L of the light direction characterises whether the light is coming from above or below, from the left or right. It is likely that we will find a bias for the light source to be located above the object, with a small bias to the left (Sun & Perona, 1998; Mamassian & Goutcher, 2001). If we call θ_L the bias on the tilt of the illuminant (with the origin corresponding to illumination from above, and positive angles to illumination from the left), and κ_S the concentration parameter about θ_L , the distribution on the light tilt direction will take the general form

$$p(\tau_L) = M(\tau_L, \kappa_S, \theta_L) = \frac{1}{2\pi I(0, \kappa_S)} e^{\kappa_S \cos(\tau_L - \theta_L)}. \quad (7)$$

The concentration parameter κ_S reflects the strength of the observer’s shading constraint (the subscript S stands for shading), namely a bias to assume that the light direction has a tilt close to θ_L .

² A measure of dispersion of the distribution around the mean is given by the *angular deviation* σ (in radians), defined as $\sigma(\kappa) = \sqrt{-2 \ln(I(1, \kappa)/I(0, \kappa))}$, where I is again the modified Bessel function. For large values of κ (sharply peaked distributions), the angular deviation converges asymptotically to the familiar standard deviation of linear statistics.

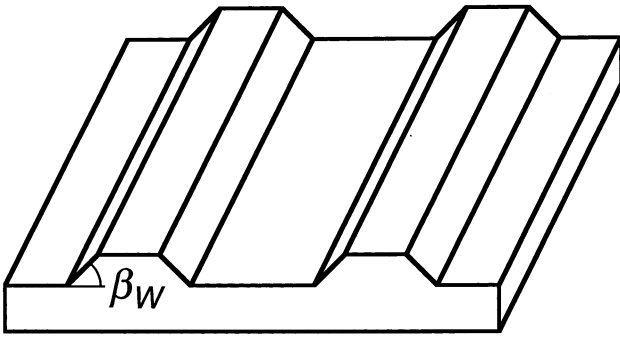


Fig. 7. Definition of the bevel angle in the world β_W .

3.2.3. Object model

The shape of the surface is obtained by protruding a planar profile (Fig. 7). Two plateaux in the profile provide the basis for the ‘narrow’ and ‘wide’ strips. The bevel between the strips has an inclination in the world that we shall call the ‘bevel angle’ β_W . The projection of the bevel angle in the image forms an angle β_P . Since the stimuli were constructed so that β_P had the same absolute value on both sides of each ridge, we assume the observer fixes β_W to have the same absolute value on both sides as well. This implies that the tilt of the surface is in the direction of the ridges, and hence:

$$\tan(\beta_P) = \tan(\beta_W) \sin(\sigma_S), \quad (8)$$

where σ_S is the slant of the surface relative to the observer. The projected bevel angle in the stimuli β_P was constant and equal to 45° . The distribution of the bevel angle itself β_W was taken to be uniform on the interval $[-90, 90]$, where positive values indicate the ‘narrow’ strips are bulging, and negative values indicate the ‘wide’ strips are bulging.

The projected width of the strips is only relevant as it provides a basis for the observer’s response of ‘narrow’ or ‘wide’. We code a potential bias to respond preferably ‘narrow’ as the proportion of objects in the world with ‘narrow’ strips bulging p_N . A value of $p_N = 0.5$ indicates no bias while a value of 1.0 indicates a completely dominant bias for ‘narrow’ strips bulging.

3.2.4. Observer model

In addition to the objects in the scene and the illumination conditions, it is essential to describe the properties of the observer. For simplicity, we model the observer as the image which is produced under orthographic projection. Under this projection system, the distance between the observer and the objects is irrelevant, only the orientation of surfaces relative to the observer is important.

Surface orientation has three degrees of freedom: slant σ_S , tilt τ_S , and roll ρ_S . The roll corresponds to a rotation about the surface normal, in this case about the surface normal of the plateaux. As we mentioned

above, the only possible value for ρ_S is so as to align the tilt direction with the ridges, as any other value will produce an image where the bright and dark strips have different widths.

We have no reason to assume any bias on the surface slant. Therefore, just as with the slant of the illuminant, the surface slant will be taken from a distribution that results in a uniform distribution in the sphere of possible normal directions (at least when the distribution of τ_S is uniform):

$$p(\sigma_S) = \sin(\sigma_S). \quad (9)$$

On the other hand, we shall allow for a potential bias on the surface tilt. Previous work has demonstrated an observer bias for the surface normal to point upwards relative to the viewing direction (Mamassian & Landy, 1998). This bias roughly corresponds to a preference for assuming that one’s viewpoint is located above the observed object. This link between surface normals and viewpoint is especially valid for objects located on a ground plane. In this case, the surface areas where the normals point downwards are mostly occluded from the observer. As with the bias for the illuminant direction, we shall use a von Mises distribution to model the bias on surface tilt, where κ_C denotes the concentration parameter:

$$p(\tau_S) = M(\tau_S, \kappa_C, \pi/2) = \frac{1}{2\pi I(0, \kappa_C)} e^{\kappa_C \cos(\tau_S - \frac{\pi}{2})}. \quad (10)$$

The concentration parameter κ_C reflects the strength of the observer’s contour constraint (the subscript C stands for contour), namely a bias to assume that the surface normals have a tilt close to 90° .

3.2.5. Bayesian combination

We now have all the elements to compute the likelihood function $p(\text{image} | \text{scene})$ which reflects the probability that an image is the projection of a scene. In our case, the scene reduces to the property that the surface has ‘narrow’ rather than ‘wide’ strips bulging (i.e., whether β_W is positive or negative). The image corresponds to the independent variables of the psychophysical experiment, that is, the combination of the shading contrast C_S , the contour contrast C_C , the projected bevel angle β_P (always 45° in the experiments) and the figure’s orientation in the frontal plane (which we shall denote θ_1). We compute the likelihood function by using the principle of total probability to include all the model parameters that apply (cf. Mamassian & Landy, 1998). Finally, using Bayes’ theorem, we compute the posterior probability (cf. Appendix A).

3.2.6. Decision rule

The last step in our modelling effort is to decide on the decision rule that will transform the posterior probability into a performance measure such as the ‘narrow

score'. Several decision rules have been used in Bayesian decision theory (cf. Yuille & Bülthoff, 1996). As discussed in the introduction however, we prefer to use here a non-committing decision rule (Mamassian & Landy, 1998). This rule is simply to set the narrow score equal to the computed posterior probability of 'narrow' given the image.

3.2.7. The full model in a nutshell

In summary, we have developed a model that uses some aspects of Bayesian analysis for our psychophysical task. The model includes four parameters: (1) a bias to prefer surfaces with 'narrow' strips bulging; (2) a bias on the tilt of the light direction; (3) a concentration parameter for the tilt of the light direction; and (4) a concentration parameter for the tilt of the surface.

We think it reasonable to assume that the bias for 'narrow' strips and the bias on the light direction are constant and independent of our experimental conditions. Recall that the concentration parameters for the light direction and the surface orientation (analogous to the inverse variance of the corresponding prior distributions) control the strength of the priors; large values of the concentration parameter result in a prior that has a strong effect on the posterior distribution and resulting

estimate. We predict that the values of the concentration parameters depend on the shading and contour contrast in a lawful way. For example, when shading contrast is high and contour contrast is low, more weight should be given to the shading prior, so that κ_S should be large relative to κ_C in this condition. Hence, in the full model there is a separate value of κ_S and κ_C for each combination of shading and contour contrast. We shall look for the best fit of the experimental data with a model that includes these parameters. We shall then look at the relationship between the concentration parameters and the shading and contour contrasts.

3.2.8. Results

The full model developed above contains 20 parameters: a bias to perceive the narrow strip as bulging (p_N), a bias to assume that light is coming from the left of vertical (θ_L), and a pair of (κ_S , κ_C) for each of the nine contrast conditions. These 20 parameters were adjusted against the 432 data points (24 orientations \times 9 contrast combinations \times 2 mirror reflections) using the downhill simplex method (Press, Teukolsky, Vetterling, & Flannery, 1992) to find the maximum likelihood fit. Two hundred million Monte Carlo simulations were used for each computation of a new simplex (cf. Appendix A). The algorithm converged in about 100 iterations.

The best fit, shown as the solid curves in Fig. 5, accounts for the trends in the data quite nicely. The fit required a slight bias to respond 'narrow' ($p_N = 0.56$). As expected, there was a small bias to assume that light is coming from the left of vertical ($\theta_L = 10.7^\circ$). This value of the lighting bias is small compared to what we have found with displays using only the shading constraint (Mamassian & Goutcher, 2001) and the range of values (from -10° to 40°) found by Sun and Perona (1998). The best-fit values of the concentration parameters are shown as the symbols in Fig. 8.

The variations of the concentration parameters κ_S and κ_C as a function of the shading and contour contrasts are as predicted. Remember that large values of the concentration parameter correspond to highly peaked prior distributions (i.e. ones with a strong effect on the resulting percept). It is therefore natural to observe an increase of κ_S as the shading contrast increases, and likewise an increase of κ_C as the contour contrast increases. The values of the concentration parameters are also of the right magnitude. The value of κ_C for the high contour contrast/low shading contrast condition is equivalent to a value of σ of approximately 1.0, which is about twice as large as the value found by Mamassian and Landy (1998) for stimuli that included only a contour cue.

Note that the values of the concentration parameters are affected by both contrasts. More precisely, κ_S increases with shading contrast but also decreases with contour contrast. Likewise, κ_C increases with contour

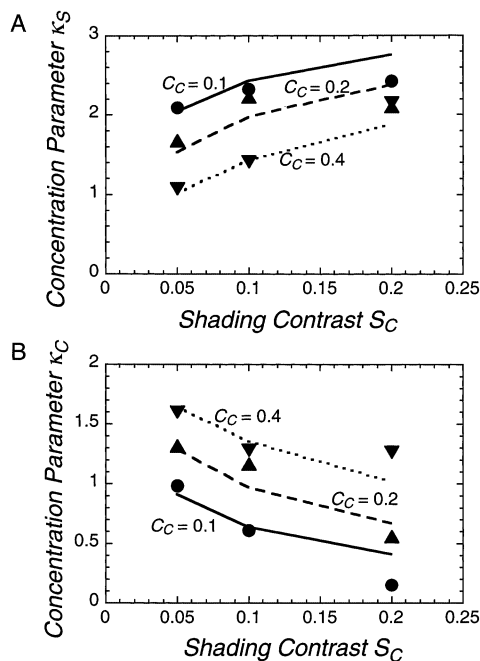


Fig. 8. Best fit distributional concentration parameters for the nine shading and contour contrast conditions. (A) The concentration parameter κ_S (effectively the weight of the shading constraint) increases with shading contrast C_S and decreases with contour contrast C_C . (B) Conversely, the concentration parameter κ_C (effectively the weight of the contour constraint) decreases with C_S and increases with C_C . The symbols correspond to the best fit of the full model. The curves in both figures correspond to the best fit of the nested model which parameterises these trends using an estimated strength for each constraint level of each constraint (see text for details).

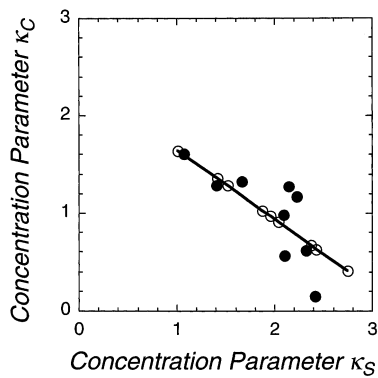


Fig. 9. Scatterplot of the values of κ_S and κ_C for each of the nine contrast combination conditions. Solid symbols are from the full model and open ones from the nested model. The strong linear relationship for the solid symbols is consistent with the nested model.

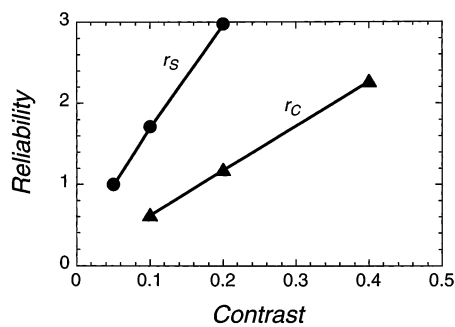


Fig. 10. Relationship between the reliability factors (r_S and r_C) and contrast. The reliability factor for the shading (resp. contour) constraint increases with shading (resp. contour) contrast. The reliability is reported in arbitrary units (the shading reliability for a shading contrast of 0.05 was arbitrarily set to one).

contrast and decreases with shading contrast. The interpretation of this double dependence is as follows. The narrowness of a prior distribution corresponds to the reliance of the observer on a particular prior. As contour reliability decreases, and hence so does κ_C , the observer must rely more on the shading cue to disambiguate the stimulus, and κ_S increases. We take advantage of this property of the concentration parameters in the nested model developed in the following section.

3.3. The nested model

We now describe a simpler model with fewer degrees of freedom to model the interdependence of κ_S and κ_C . From the discussion above of the results obtained with the full model, it appears that κ_S and κ_C are a function of both shading and contour contrast. We do not know how the concentration parameters κ_S and κ_C vary with contrast. However, it is clear from the full model discussed above that increasing the shading contrast leads to higher κ_S (and likewise for the contour contrast and κ_C). Thus, for each constraint and contrast, we

include as a parameter a ‘reliability factor’ $r_{\text{constraint}}(\text{contrast})$. Then, the values of κ_S and κ_C will depend on the relative reliabilities in each condition as follows:

$$\begin{aligned}\kappa_S &= s_S r_S(C_S) / (r_S(C_S) + r_C(C_C)) \\ \kappa_C &= s_C r_C(C_C) / (r_S(C_S) + r_C(C_C)),\end{aligned}\quad (11)$$

where $s_{\text{constraint}}$ is a scale factor for that particular constraint (shading or contour) and the $r_{\text{constraint}}$ values are the reliabilities. This model contains ten degrees of freedom: six reliability factors, two scale factors, plus p_N and θ_L . In fact, because all the reliability factors are used as ratios, there is one degree of freedom wasted. We can therefore arbitrarily set $r_S(0.05) = 1$ and end up with a final model with only nine degrees of freedom.

A straightforward prediction can be obtained from this nested model:

$$(\kappa_S/s_S) + (\kappa_C/s_C) = 1. \quad (12)$$

In other words, κ_S and κ_C should be linearly related. The relationship between the values of κ_S and κ_C derived from the fit of the full model in the previous section is shown in Fig. 9. A linear regression results in a correlation coefficient equal to -0.76 . This high correlation is consistent with the simpler, nested model. We therefore repeated the fitting procedure (downhill simplex method with 200 million samples for the Monte Carlo simulations) for the nine-parameter model. The fit required similar biases for ‘narrow’ percepts ($p_N = 0.56$) and for a light coming from the left of vertical ($\theta_L = 11.8^\circ$).

The scale factors were 3.33 and 2.36 for the shading and contour cues, respectively. The fit values of κ_S and κ_C are shown by the curves in Fig. 8. This model fits the data almost as well as the full model and nicely summarises the effects of stimulus contrast manipulations. These two models have different numbers of parameters. Since the models were fit using a maximum likelihood criterion, the quality of the fits may be compared using a nested hypothesis test (Mood, Graybill, & Boes, 1974). Unfortunately, this test rejects the hypothesis that the nested, nine-parameter model fits the data as well as the full, 20-parameter model ($\chi^2(11) = 46$, $P < 0.001$).

The advantage of the nested model, however, is that it allows us to compare the overall effects of the shading and contour constraints. In particular, the scale factors reflect the strength of each constraint (this can be seen by setting the reliability factor of the other constraint to zero). The fact that the scale factor for shading was larger than the one for contour indicates that the shading constraint had more weight in the final decision of the observer. The reliability factors map stimulus contrasts to concentration parameters. As expected, they increase with increasing contrast (Fig. 10).

The relationship is quite linear over the range studied, suggesting a further nested model that parameterizes these relationships as straight lines passing through the origin (and saves four more parameters).

3.4. *Intermediate discussion*

We described the full and nested models as utilising Bayesian calculations, but they certainly are not true Bayesian decision makers. There are three fundamental ways these models depart from more traditional Bayesian models. First, the likelihood function was deterministic and binary rather than stochastic and continuous. In general terms, the likelihood function characterises the mapping between the distal and the proximal stimulus (the 3D world and its 2D projection, respectively). When the image is complex or when the contrast is low, the perception of the proximal stimulus is limited by external and internal sources of noise (e.g. added pixel noise or noise in the response of neurons). Because of the noise, the same distal stimulus can produce a distribution of proximal stimuli, with some images occurring more frequently than others. In our task, noise will affect the mapping of world states (fixed values of bevel angle, albedos, surface and illumination directions) to image attributes (stimulus contrasts, orientation and projected bevel angle). In the models we have described so far, we have disregarded all sources of noise. We argued that with our reasonably high-contrast, easily visible stimuli, quantum fluctuations in the light incident at the cornea would result in very little variability in the estimates of the world states. If the external noise is non-significant, then the only way to put the noise back into the stimulus, as it were, is to assume that it is the observer that is noisy. That is, we would have to model an observer whose estimation of stimulus contrasts and angles was more variable than photon noise would require (see the internal noise model described in the next section). This approach has been taken by several authors who generally used control experiments or extant literature from simple discrimination experiments to gauge the variability of observer estimates of the relevant stimulus variables (see, e.g. Eagle & Blake, 1995; Crowell & Banks, 1996). It was our intuition that the variability of the various stimulus attributes that are used by the shading and contour cues (i.e. image contrasts and image contour orientations) was moderate. Assuming that the internal noise of the observer was as small as the external noise, we were led to build a likelihood function that was both deterministic and binary: either the image was consistent with a given description of the world or it was not.

The second non-Bayesian heresy in our modelling is the way we treat the priors. We have already stressed that the ambiguity inherent to all proximal stimuli can only be circumvented thanks to prior constraints.

Therefore, it seems natural to think of priors as sources of information that are independent of the stimulus. In our model, we have represented prior distributions as functions of two parameters, the mean and the variance of a probability distribution. While the mean was indeed independent of the stimulus, we allowed the variance to change according to the stimulus contrast. We argued that the variance characterized the reliability of the prior, and as such should be estimated from the stimulus properties. This argument is similar in spirit to the dynamic re-weighting of depth cues advocated by Landy et al. (1995).

The final departure from traditional Bayesian modelling deals with our choice of the decision rule. A true Bayesian modeller would first choose a gain function that rewards maximally the ‘correct’ response (the one that matches the modelled world state that gave rise to the image) and gradually less the ‘incorrect’ ones. The Bayesian modeller would then use as a decision rule the one that maximizes the expected gain. For each choice of a gain function, there is only one Bayes’ decision rule. For instance, if the gain function is binary and rewards all incorrect responses equally, then the Bayes’ decision rule is the commonly-used maximum a posteriori (MAP) rule. One important characteristic of all Bayes’ decision rules is that they are deterministic. As a consequence, the same proximal stimulus should lead to the same response if there is no noise in the system. We argued above that the internal noise levels necessary to explain the observer variability were probably too large to be realistic. As a result, we were required to derive the variability of observers’ responses at the level of the decision rule. We chose to use a non-committing decision rule for which observers responded that the stimulus was ‘narrow’ on a proportion of trials equal to the estimated posterior probability that the narrow strips were bulging. While Bayes’ decision rules are optimal, the non-committing rule is sub-optimal but arguably more realistic from a psychological point of view, as it allows the observer to essentially ‘sample’ the environment to be able to notice when assumptions about the world, coded as priors, are no longer valid (Mamassian, Landy, & Maloney, 2001).

In summary, the first two models we have presented are Bayesian in spirit, but deviate from a typical Bayesian model in several fundamental ways. We can try to justify these modelling choices in three ways. First, one can think of this as merely a descriptive model (what observers do) of human behavior rather than a normative one (what observers should do), and that this descriptive model uses (or misuses) Bayesian calculations as a foundation. Thus, since observers do not know the statistics of the world, they use their best guess to constrain the solution from the cues they have. In an ad hoc fashion, observers examine the cues present, estimate their relative reliabilities, and use

those to play off the relative influence of the cues at hand, much as they do when combining multiple cues that provide independent quantitative depth information (Landy et al., 1995). This heuristic description of what we have done gives a context for understanding why a change in contour contrast would change κ_S . There is no reason a change in image contour contrast should affect the reliability of shape-from-shading (which does not use those contours). Instead, this heuristic uses relative reliability to determine constraint strength much as relative reliability determines relative weight for cue combination.

A second take on what we have done is that it is an approximation to a calculation that does not violate the rules of Bayesian modelling so severely. In particular, consider the priors in our model that are parameterised by κ_S and κ_C , which are in turn dependent on the image itself. Thus, as mentioned above, the ‘prior distributions’ in the model are not, in fact, fixed prior to observing the input image. However, one might be able to fix this problem by considering more complex prior distributions.

Suppose, for example, that we reparameterise the model of surface reflectance as an average reflectance $r_A = (r_B + r_D)/2$, and reflectance contour contrast $r_C = (r_B - r_D)/(r_B + r_D)$. Finally, we no longer assume that the prior distributions completely factor (i.e. not all parameters are independent). In particular, assume that $p(\tau_S | r_C)$ is a von Mises distribution, but that its concentration parameter κ_C depends on the contour contrast r_C (this is related to the ‘competitive priors’ described by Yuille & Bülthoff, 1996). This is within the standard Bayesian formulation, as the prior on τ_S can validly depend on r_C , since this is not an image measurement but rather a property of the world. Since image contour contrast C_C may be used to estimate r_C , a given image contour contrast can effectively select a distinct marginal of the joint prior $p(\tau_S, r_C)$. This avoids the dependence of the contour prior on the input image. Unfortunately, a similar trick can not be used for the shading prior, as the image shading contrast depends on both the illumination contrast (the relative strength of ambient and point illumination) and on the object geometry (surface normal and illuminant direction). And, even if that problem were circumvented, this formulation does not include the dependency of κ_C on illumination contrast, or of κ_S on reflectance contour contrast, even though the data demand such a dependency.

The third possibility is to formulate a new model that has fixed priors, and for which the effects of varying image contour and shading contrast result from noise in the estimation of the various image measurements. We discuss just such a model next.

3.5. The internal noise model

In the first two models described above, shading and contour contrasts had a direct influence on the reliability of the shading and contour constraints. Even though one could conceive that these constraints are first quickly updated before being used for further processing, the idea to have variable prior constraints remains counter-intuitive. Alternatively, we can let the shading and contour contrasts influence the likelihood function rather than the priors. In particular, if each image measurement is affected by noise internal to the system, we can expect this noise to be dependent on image contrast. We now describe a third model that attempts to take into account the effects of shading and contour contrasts as resulting from different levels of internal noise in the visual system.

Internal noise makes the estimation of image attributes non-deterministic. Since we have modelled the image as a list of four variables, we can expect to have four associated sources of internal noises. However, it turned out that it was only useful to add noise to the shading contrast (C_S), and not to the other three variables (C_C , β_P , θ_I). The reflectance of the contours was set arbitrarily so C_C does not play a critical role. Simulations showed that internal noise added to β_P and θ_I only made the decisions more random (choice probabilities closer to 0.5), and did not shift the transition points between narrow and wide.

Because the shading contrast C_S varies between zero and one, we cannot use a normal distribution to model internal noise. Instead, we chose to use the beta distribution. Like the normal, the beta distribution has two degrees of freedom: the mean and the standard deviation.³ We assume estimation of C_S is unbiased; we fix the mean of the beta distribution to C_S (i.e. to 0.05, 0.1, or 0.2). We allow the standard deviation of the beta distribution to vary between the nine conditions of the experiment.

In addition to the nine degrees of freedom corresponding to the standard deviation of the noise distribution, the internal noise model has four other degrees of freedom that are similar to the ones in the full model: (i) a narrow bias, (ii) a bias to the left for the light source position, (iii) the concentration parameter for the shading prior distribution, and (iv) the concentration parameter for the contour prior distribution.

³ The beta distribution is a function restricted to the domain $[0, 1]$ that approximates a normal distribution when its mean is far away from both 0 and 1. It is usually defined by the two degrees of freedom p and q : $\text{beta}(p, q)(x) = a(p, q)x^{p-1}(1-x)^{q-1}$; for all x in $[0, 1]$, where $a(p, q) = \Gamma(p+q)/(\Gamma(p)\Gamma(q))$; for $p > 0$ and $q > 0$. When certain conditions are satisfied, the two degrees of freedom can be exchanged with the mean and standard deviation of the distribution by inverting the following relations: mean = $p/(p+q)$, variance = $pq/((p+q+1)(p+q)^2)$.

These latter two parameters are now fixed across the nine conditions. The best fit of this model was again estimated using the downhill simplex method to find the maximum likelihood fit. Because the current model focuses on the shading contrast, we allowed for finer effects of the shading contrast by increasing its resolution to 200 bins of size 0.01. The decision rule was again set to the non-committing decision rule described above. Because of the increased resolution of the shading contrast, we increased the number of Monte Carlo simulations to one billion (10^9) for each computation of a new simplex (each new iteration took about 5 h to compute on a UNIX workstation). The algorithm converged in about 100 iterations (i.e. 3 weeks).

The best fitted values of the four parameters that were kept constant across all nine conditions were as follows: (i) bias to respond ‘narrow’: $p_N = 0.57$; (ii) bias to assume that light was coming from the left of vertical: $\theta_L = 12.6^\circ$; (iii) concentration parameter for the shading cue: $\kappa_S = 2.62$; (iv) concentration parameter for the contour cue: $\kappa_C = 1.03$. In agreement with the nested model, we note that the concentration parameter for the shading cue was larger than that for the contour cue, suggesting that the prior on the illumination was stronger than the prior on viewpoint for this particular experiment.

The estimated standard deviations of the shading contrast noise distributions are summarised in Fig. 11. The noise increases with shading contrast consis-

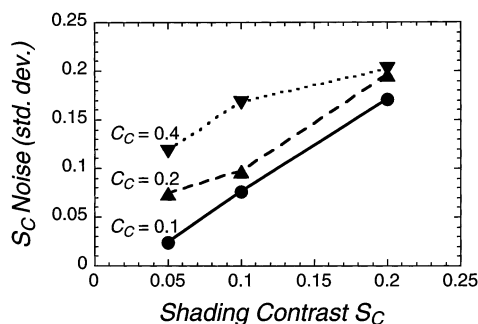


Fig. 11. Estimated internal noise strength. The estimated standard deviation of the internal shading contrast noise is plotted as a function of the shading contrast C_S and contour contrast C_C .

Table 1
Goodness-of-fit of the three models of the interaction of shape from shading and contour

Model	Degrees of freedom	Log(likelihood) of best fit
Full model	20	−4091
Nested model	9	−4114
Internal noise model	13	−4385

tent with Weber’s law. In addition, the noise increases with contour contrast, a result that is less intuitive.

The performance of the three models can be compared in the Table 1 where smaller absolute values of the log-likelihood indicate better fits. From this table, we can see that the model based on internal noise fits the data substantially less well than the nested model, even though it has more degrees of freedom. Since the internal noise model is not a sub-case of the other two models, the nested hypothesis test does not apply.

4. General discussion

We reported the results of a psychophysical experiment that shows a clear interaction of two visual constraints in the interpretation of ambiguous stimuli. We then proposed models that captured the essence of human performance. The models were inspired by Bayesian decision models but departed significantly from these. We next discuss one alternative approach, as well as implications of these results for everyday perception.

4.1. Strength models

A more traditional approach in modelling psychological behaviors is to use strength-based, or Thurstonian models. If such a model were to be used here, one would have to choose a set of parametric functions to describe the contribution of every constraint for the interpretation of the stimuli. Each parametric function would characterise the strength of a single constraint assuming that the others made no contribution. Higher strengths would correspond to stronger evidence for the ‘narrow’ percept. Typically, the mapping of strength to response can be modelled using a standard normal distribution (corresponding to Thurstone’s case V). Then, the interaction of the two constraints would be modelled as a weighted combination of the strengths of the individual constraints. In a study on the interaction of multiple depth cues, Doshier, Sperling, and Wurst (1986) used exactly such a model to fit data on the disambiguation of Necker cube stimuli. An additive strength model fit their data extremely well (in fact, as well as binomial variability would allow).

Even though our results could certainly be modelled by a strength model, we believe that our approach offers some important advantages. First, our model was built following a principled way based on physical properties (geometry of 3D to 2D mapping, illumination models) and psychological components (prior constraints, decision rules). As a result, the combination rule of the two constraints was not imposed by the modeller (using, for instance, a linear

combination) but instead naturally emerged from the modelling effort. This allowed us to study the interaction a posteriori with the help of our nested model. A second advantage of our approach resides in the interpretation of the parameters of our model. Because there are no ad hoc components in the model, each parameter has a clear meaning. This property should allow us to provide quantitative predictions for future studies.

4.2. Scope of the study

One critical aspect of our psychophysical experiment was the focus on the very first interpretation of the ambiguous stimuli. We reasoned that if visual constraints had any effect on human performance, this effect would appear at the beginning of the stimulus processing rather than later on. In our previous study (Mamassian & Landy, 1998) and the present one, we found that visual constraints indeed had a significant effect on the first interpretation of the stimuli. Our experimental conditions differ somewhat from everyday perceptual experience in that objects in our environment do not disappear suddenly from sight after a fraction of second. It is well known that when an ambiguous stimulus is seen for a long time, the interpretation starts to alternate back and forth between the different alternatives. We would predict that the time periods of seeing one particular interpretation is proportional to the posterior probability for this interpretation. In other words, we believe that the prior constraints are not only useful for the very first interpretation of a stimulus but instead remain useful as long as the stimulus is observed.

Even though our models were applied to only one type of stimuli, we would like to believe that our results could be extended to other contexts. Our models provide quantitative estimates of the shading and contour constraints, in terms of both bias and reliability. These estimates can therefore be used in future studies to predict the effect of illumination and viewpoint in different experimental conditions.

4.3. Summary

In this paper, we have described a psychophysical experiment whose purpose was to investigate how human observers use multiple prior constraints to interpret ambiguous images. We found that the interpretation of an ambiguous stimulus chosen by observers was effectively a compromise (in a probabilistic sense) between that indicated by the two constraints governing the two depth cues present in the stimulus. When the stimulus information for one cue was more reliable than for another (i.e. had higher stimulus contrast), then the corresponding constraint took precedence.

Acknowledgements

We would like to thank Laurence Maloney for insightful discussions at different stages of this project as well as Marty Banks, David Knill and two anonymous reviewers for helpful comments. Funding was provided by NIH grant EY08266, AFOSR grant 93NL366 and the Human Frontier Science Program grant RG0109/1999-B.

Appendix A

This appendix shows the derivation of the posterior probability from the knowledge of the prior distributions described in the text. The posterior probability is

$$p(\text{narrow} \mid \text{image}) = \int_0^{\pi/2} p(\beta_w \mid C_C, C_S, \beta_P, \theta_I) d\beta_w, \quad (13)$$

By the principle of total probability, the integral may be expanded to

$$\int_{\Omega} p(\beta_w, \sigma_S, \tau_S, \sigma_I, \tau_I, r_B, r_D, I_a, I_p \mid C_C, C_S, \beta_P, \theta_I) \times d\beta_w d\sigma_S d\tau_S d\sigma_I d\tau_I dr_B dr_D dI_a dI_p, \quad (14)$$

where Ω is the domain of integration, which is the entire domain of most of the variables, but only positive values of β_w , corresponding to narrow strips bulging. By Bayes' rule the integrand may be rewritten as a ratio of three terms

$$\frac{p(C_C, C_S, \beta_P, \theta_I \mid \beta_w, \sigma_S, \tau_S, \sigma_I, \tau_I, r_B, r_D, I_a, I_p) \times p(\beta_w, \sigma_S, \tau_S, \sigma_I, \tau_I, r_B, r_D, I_a, I_p)}{p(C_C, C_S, \beta_P, \theta_I)} \quad (15)$$

The first term in the numerator corresponds to the likelihood that a set of world parameters produces an image. It takes the value 'one' if the world parameters correspond exactly to the image and 'zero' otherwise. By assuming independence of our priors, the second term in the numerator may be rewritten as

$$p(\beta_w)p(\sigma_S)p(\tau_S)p(\sigma_I)p(\tau_I)p(r_B)p(r_D)p(I_a)p(I_p). \quad (16)$$

These prior probabilities were detailed in the sections above. Finally, the denominator can be simply computed using the fact that it is the normalising factor such that

$$p(\text{narrow} \mid \text{image}) + p(\text{wide} \mid \text{image}) = 1. \quad (17)$$

Unfortunately, the calculation of the posterior distribution proved to be impossible in closed form. Hence, we resorted to numerical methods. We used a Monte Carlo simulation to compute the likelihood values (the values of $p(\text{image} \mid \text{narrow})$ and $p(\text{image} \mid \text{wide})$). Given a set of distributional parameters (p_N , κ_C , κ_S and θ_L), we repeatedly generated scenes based on the prior

distributions, resulting in sample values of all scene parameters (β_w , σ_s , τ_s , σ_l , τ_l , r_B , r_D , I_a and I_p). If these were consistent with the image (C_s , C_c , β_p , θ_l), then the sample value was saved. For this purpose, the contrasts were binned in 100 intervals of size 0.02 (contrasts vary between -1 and 1), the projected bevel angle (β_p) had bins of width 10° centered on the only two possible values of $\pm 45^\circ$, and the orientation of the image θ_l was binned in 24 intervals of size 15° (θ_l varies between -180° and 180°). The binning resolution had only small effects on the model's performance; the final binning values were chosen to reduce the computation time. We generated 200,000,000 scenes and kept a count of the number of samples with positive ('narrow') and negative ('wide') values of β_w . The posterior probability was then computed as

$$p(\text{narrow} \mid \text{image}) = \frac{p_N \times p(\text{image} \mid \text{narrow})}{p_N \times p(\text{image} \mid \text{narrow}) + (1 - p_N) \times p(\text{image} \mid \text{wide})} \quad (18)$$

Finally, the values of the parameters were varied so as to provide the best fit, in the maximum likelihood sense, of the resulting posterior probabilities to the observed data.

References

- Batschelet, E. (1981). *Circular statistics in biology*. London, UK: Academic Press.
- Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: a new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments and Computers*, 25, 257–271.
- Crowell, J. A., & Banks, M. S. (1996). Ideal observer for heading judgments. *Vision Research*, 36, 471–490.
- Doshier, B. A., Sperling, G., & Wurst, S. A. (1986). Tradeoffs between stereopsis and proximity luminance covariance as determinants of perceived 3D structure. *Vision Research*, 26, 973–990.
- Eagle, R. A., & Blake, A. (1995). Two-dimensional constraints on three-dimensional structure from motion tasks. *Vision Research*, 35, 2927–2941.
- Foley, J. D., van Dam, A., Feiner, S. K., & Hughes, J. F. (1996). *Computer graphics: principles and practice* (2nd ed.). Reading, MA: Addison-Wesley.
- Gregory, R. L. (1980). Perceptions as hypotheses. *Philosophical Transactions of the Royal Society of London, Series B*, 290, 181–197.
- Hogervorst, M. A., & Eagle, R. A. (1998). Biases in three-dimensional structure-from-motion arise from noise in the early visual system. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, 265, 1587–1593.
- Horn, B. K. P. (1990). *Robot vision*. Cambridge, MA: MIT Press.
- Kersten, D. (1990). Statistical limits to image understanding. In C. Blakemore, *Vision: coding and efficiency* (pp. 32–44). Cambridge, UK: Cambridge University Press.
- Knill, D. C., & Richards, W. (1996). *Perception as Bayesian inference*. Cambridge, UK: Cambridge University Press.
- Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. (1995). Measurement and modelling of depth cue combination: in defense of weak fusion. *Vision Research*, 35, 389–412.
- Maloney, L. T. (2001). Statistical decision theory and biological vision. In D. Heyer & R. Mausfeld, *Perception theory: conceptual issues*. New York: Wiley.
- Mamassian, P., & Goutcher, R. (2001). Prior knowledge on the illumination position. *Cognition*, 81, B1–B9.
- Mamassian, P., & Kersten, D. (1996). Illumination, shading and the perception of local orientation. *Vision Research*, 36, 2351–2367.
- Mamassian, P., & Landy, M. S. (1998). Observer biases in the 3D interpretation of line drawings. *Vision Research*, 38, 2817–2832.
- Mamassian, P., Landy, M. S., & Maloney, L. T. (2001). Bayesian modelling of visual perception. In R. Rao, B. Olshausen, & M. Lewicki, *Probabilistic models of the brain: perception and neural function*. Cambridge, MA: MIT Press.
- Marr, D. (1982). *Vision*. San Francisco, CA: W.H. Freeman.
- Mood, A. M., Graybill, F. A., & Boes, D. C. (1974). *Introduction to the theory of statistics*. New York, NY: McGraw Hill.
- Poggio, T., Torre, V., & Koch, C. (1985). Computational vision and regularization theory. *Nature*, 317, 314–319.
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (1992). *Numerical recipes in C* (2nd ed.). Cambridge, UK: Cambridge University Press.
- Ramachandran, V. S. (1988). Perception of shape from shading. *Nature*, 331, 163–166.
- Rock, I. (1983). *The logic of perception*. Cambridge, MA: MIT Press.
- Stevens, K. A. (1981). The visual interpretation of surface contours. *Artificial Intelligence*, 17, 47–73.
- Sun, J., & Perona, P. (1998). Where is the sun? *Nature Neuroscience*, 1, 183–184.
- Ullman, S. (1979). *The interpretation of visual motion*. Cambridge, MA: MIT Press.
- Weiss, Y., & Adelson, E. (1998). Slow and smooth: a Bayesian theory for the combination of local motion signals in human vision. MIT AI Lab. Technical Report Number 1624.
- Yuille, A., & Bülthoff, H. H. (1996). Bayesian decision theory and psychophysics. In D. C. Knill, & W. Richards, *Perception as Bayesian inference* (pp. 123–161). Cambridge, UK: Cambridge University Press.